

NATURAL-LANGUAGE VOICE-ACTIVATED PERSONAL ASSISTANT

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The application claims benefit of U.S. Provisional Patent Application No. 60/236,650, filed September 29, 2000, and entitled "NATURAL-LANGUAGE VOICE-ACTIVATED PERSONAL ASSISTANT," which is hereby incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] The present invention relates generally to personal assistant and more particularly to natural-language voice-activated personal assistant.

[0003] For the past few years, personal assistants have been growing at a phenomenal rate. Several companies, including Palm, Inc. and Handspring, Inc., have successfully entered the market. Their products enhance personal productivity to a certain degree. These personal assistants are computing devices and thus also referred to as "digital assistants." However, such assistants have weaknesses.

[0004] Typically, such assistants are handheld, meaning that they can be held in one's hand or put in a shirt pocket. As time goes by, handheld can even imply a little badge on one's shirt, which sometimes can be referred to as an Internet appliance or wearable computer. These assistants are typically quite small with very small keyboards. As a result, it is not only tedious but also time consuming to interact with such assistants.

[0005] One way to attempt to get around the interaction problem is to include a voice-recognition mechanism in the assistant. IBM recently announced that they are contemplating incorporating voice-recognition mechanisms into personal assistants. There are also cellular phones where you can use key words verbally to access phone numbers. However, the key words have to match the names you previously stored in the phone.

[0006] Voice-recognition alone is insufficient to solve the interaction problem. Humans can express an idea in so many different ways. For example, after recording his meeting time with Alice into the calendar of an assistant, Joe wants to find out the time of the meeting. He can ask for the information in many ways. Joe may ask, "When is my meeting time with Alice?", "When should I be meeting Alice?", "Damn! Should I be meeting Alice tomorrow?" or "Tell me when is Alice meeting me." Clearly, voice-recognition alone is not sufficient to resolve the different ways of expression and retrieve Joe's meeting time with Alice.

[0007] Another issue involved is that human expression can be ambiguous in other ways. For example, Joe can ask, "Meeting Alice?" or "When to meet Alice?". Such expressions are ambiguous and, strictly speaking, are syntactically incorrect. But the assistant should be able to retrieve the meeting time for Joe, just like a good human assistant can. Such different or ambiguous expressions are common in everyday conversation, and should be expected when one is using his personal assistant to get phone numbers, retrieving to-do list or looking up calendar events.

[0008] It should be apparent from the foregoing that voice-recognition software alone is insufficient to make personal assistants or cell phones applicable for common everyday expressions.

SUMMARY OF THE INVENTION

[0009] The present invention provides a personal assistant that understands human expressions to retrieve information for a person. In one embodiment, the personal assistant is a handheld computing device, which can also be referred to as a "digital assistant." For example, James can be asking for David's phone number in many different ways, such as "Let me have David's phone number.", "What is David's phone number?" or "David's phone number, please." Through the present invention, the personal assistant can still extract the phone number for David. If there is more than one David in the address book, the assistant can ask James to resolve the ambiguity. For example, the assistant can ask James,

"Are you asking for David Chaos or David Tsunami?" Depending on the response, the assistant can access the phone number.

[0010] In one embodiment, a handheld personal assistant includes a voice-recognizer and a natural-language processor. The voice-recognizer can transform an expression received from a person (i.e., user) into a different mode of information. This mode can be text or other non-waveform representation.

[0011] In another embodiment, the recognizer can be previously trained to recognize the person's voice, but not another person's voice. Based on the person's voice, the assistant can only allow the person to access information that is personal to the person.

[0012] The natural-language processor can process the mode of information to extract, from a database, a piece of information that is personal to the person. The natural-language processor can still extract the piece of information even when the person declares the expression differently, or even if the expression is ambiguous. If the assistant cannot resolve an ambiguity in the expression, the assistant can provide the person with a number of alternatives to resolve the ambiguity or otherwise ask or seek clarification.

[0013] In processing, the natural-language processor can analyze the expression grammatically and semantically to transform at least a part of the expression into at least one instruction.

[0014] The piece of information can be a personal address book, a to-do-list or a calendar. The piece of information can depend on the context under which the person made the expression, and the expression can be just one word.

[0015] The handheld personal assistant can also include a display to display the piece of information. In another embodiment, the assistant can include a voice synthesizer, or more commonly known as a speech synthesizer, to transform the piece of information into sound to communicate to the person.

[0016] In one embodiment, the piece of information was entered into the assistant by the user. The piece of information can be entered through use of

a keypad or keyboard or through voice. The assistant can further include a categorizer that stores the piece of information into the database. To assist the categorizer, the person can identify a category.

[0017] In another embodiment, the assistant has a receiver to receive expressions and a transmitter to transmit information to a second system. This transmission can be done wirelessly. The expressions can be transformed into a different mode of information either by the assistant, by the second system, or a combination of the two. It would then be up to the second system to process the mode of information to extract, from a database, a piece of information that is personal to the person. After extraction, the second system transmits the piece of information back to the handheld personal assistant.

[0018] Other aspects and advantages of the present invention will become apparent from the following detailed description, which, when taken in conjunction with the accompanying drawings, illustrates by way of example the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] The present invention will be readily understood by the following detailed description in conjunction with the accompanying drawings, wherein like reference numerals designate like structural elements, and in which:

FIG. 1 is a block diagram of a personal digital assistant according to one embodiment of the invention; and

FIG. 2 illustrates a language processing system according to one embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0020] One embodiment of the invention includes a handheld personal assistant with a voice recognizer and a natural-language processor. In the following description, the voice recognizer is user-dependent; however, the invention can also be applicable to a user-independent recognizer.

[0021] After getting the handheld personal assistant, a person (i.e., the user) trains the voice recognizer to recognize his voice. This can be done through reading a piece of predetermined text to the assistant, as in IBM's ViaVoice product line. After reading the piece of text, the assistant is programmed to recognize the user's voice. For some products in the market place, the accuracy of voice recognition can then be more than 95%.

[0022] After the training, the recognizer will be geared to recognize the user's voice, but not another person's voice. The user can then, for example, ask for the time of his meeting with Joe tomorrow. The voice-recognizer can transform this expression into a different mode of information, for example, text. Based on the person's voice, the assistant can allow the user to access the piece of information that is personal to him.

[0023] In one embodiment, the assistant obtains a context from the user query. This can either be a direct context request or an inferred context. For example, the user can directly say one of the following three options: Calendar, Address Book, or To-do list. Alternatively, the assistant can determine context based on the query. For example, a request for a phone number would place the assistant into the address book context. The assistant will prompt the user for clarification if a context is ambiguous.

[0024] Once a context is established, the context is maintained through the session or until the user requests another context, either directly or indirectly. For example, if the user requests the following: "What is the phone number for Joe?", this will place the assistant into Address Book context. If he follows up with "Tom?", the assistant will retain the most recent context (Address Book, a request for phone number) and apply it to the search for Tom. As an example of indirectly changing context, again if the user asks the phone number of Joe, after getting the response, the user can subsequently asks, "Any

meeting with him?" This can switch the context for the assistant from phone book to calendar, while keeping certain relevant information in the original question, in this case, "Joe". The two questions can be in the same session, and "him" in the second question is replaced by the word, "Joe", to get the answer for the question.

[0025] The natural-language processor can process the mode of information to extract, from a database, a piece of information that is personal to the person. In one embodiment, the assistant can use some of the natural-language processing methods described in U.S. Patent Application Nos.: 09/347,184, filed July 2, 1999 (now U.S. Patent No. _____) and 09/387,932, filed September 1, 1999, which are incorporated herein by reference.

[0026] In processing, the processor can analyze the expression grammatically and semantically to transform at least a part of the expression into at least one instruction. In one embodiment, the instruction can be a query to a database. For example, the processor transforms the expression into a SQL query to search for information in the database.

[0027] In another embodiment, the processor transforms the expression into question structures and question formats. The question formats can then be transformed into instructions to retrieve information. To explain in more detail, each phrase in the expression can be linked to a category, and the categorized representation of the expression can be known as a question structure. In other words, a question structure can be a list of categories. For example, the question structures of the expression, "cash cow" can be "finance" "animals" and just "finance" if the expression is considered as one single phrase. In this example, the expression, "Cash cow?" is linked to two question structures. After the question structures representing the expression have been selected, one or more question formats can be identified. Each question format can be a pre-defined question with one or more phrases and one or more categories. The following can be a question format:

[0028] What is "name" "phone number"?

[0029] The expression, "What is Joe's phone number?", falls under the above question format.

[0030] Each category in a question format has a number of corresponding phrases. For example, the corresponding phrases of the category "name" can include all of the names in the directory of the database. After the identification of the question format, the processor can transform it into instructions. In one situation, the instruction can be database queries.

[0031] The processor can still extract the piece of information even when the user declares the expression differently, or even if the expression is ambiguous. If the assistant cannot resolve an ambiguity in the expression, the assistant can provide the user with a number of alternatives to resolve the ambiguity. For example, the user may be given the alternatives to pick between Joe Smith and Joe Winter.

[0032] The piece of information can be associated with a personal address book, a to-do-list or a calendar. The piece of information can depend on the context under which the person made the expression, and the expression can be multiple words or just one word.

[0033] The handheld personal assistant can also include a display to display the piece of information. In another embodiment, the assistant can include a voice synthesizer, or more commonly known as a speech synthesizer, to transform the piece of information into sound to communicate to the person.

[0034] In one embodiment, the piece of information was entered into the assistant by the user. The piece of information can be entered through use of a keypad or keyboard or through voice. The assistant further includes a categorizer that stores the piece of information into appropriate areas in the database. To assist the categorizer, the person can identify a category. This can be done, for example, by declaring the following, "Entering new information into the address book." The natural-language processor will know that the next expression is particularly for entering information into the address book. The user can then declare, "Joe Montani's phone number is 650-1234567." As explained in U.S. Patent Application 09/496,863, filed February 2, 2000, which is hereby incorporated by reference , after identifying the category an input belongs to, the input can be linked to that category.

Similarly, in the present situation, the categories can be the "name" and the "phone number" categories. After identifying both of them by the natural-language processor, the categorizer places "Joe Montani" into the "name" category, and "650" into the "area code" section and "1234567" into the "main phone number" section of the "phone" category. As in requesting for information, selecting context for entry can also be done indirectly. For example, the user declares, "Joe's phone number is 1234567. Then he states, "Meeting with him at 5pm tomorrow." The context for the second piece of information can be switched from address book to calendar. Again certain information in the first entry can be retained for the second entry to resolve ambiguity in the second entry, which in this case is the name, "Joe".

[0035] Input of information and retrieval of information can be considered contexts, therefore, they can be switched like other contexts, either by direct requests or inferred from the question. In one embodiment, after the user has entered the information, he can explicitly request an end of the context by declaring, for example, "I have finished entering new information." In another embodiment, the user may indirectly switch context by making a query, thus switching to the information retrieval context. Later, when the user needs Joe Montani's phone number, he can ask, "Let me have Joe Montani's phone number". In another embodiment, he can set the context of "address book" before asking for Joe Montani's phone number.

[0036] FIG. 1 is a block diagram of a personal digital assistant (e.g., handheld personal assistant) 100 according to one embodiment of the invention. The PDA 100 includes conventional PDA hardware 102 that is typically found within a PDA. The PDA hardware 102, for example, includes a processor, RAM, ROM, operating software, a display (optional), and a wireless modem (optional). The PDA 100 also includes a microphone 104, a voice recognizer 106 and a natural language processor 108. The microphone 104 receives a voice input (e.g., user query) which is supplied to the voice recognizer 106. The voice recognizer 106 delivers the voice input to the natural language processor 108. The natural language processor 108 processes the voice input to understand the voice input in a natural language context. The natural language result can then be supplied to the PDA

hardware 102. This enables the voice input to be processed and thus understood using the voice recognizer 106 and the natural language processor 108. Once understood, the input can be used to direct the PDA hardware 102 to perform predetermined actions (e.g., perform an action, retrieve content, launch an application, etc.).

[0037] In another embodiment, the assistant has a receiver to receive expressions and a transmitter to transmit information to a second system. This transmission can be done in a wired or wireless manner. The expressions can be transformed into a different mode of information either by the assistant, by the second system, or a combination of the two. It would then be up to the second system to transform the expression into a different mode of information and to process the mode of information to extract, from a database, a piece of information that is personal to the person. After extraction, the second system transmits the piece of information back to the handheld personal assistant. As an example, the second system can be a server computer and the assistant can be a client computer.

[0038] FIG. 2 illustrates a language processing system 200 according to one embodiment of the invention. The language processing system 200 includes a network 202, PDA 204, PDA 206, and a language processing server 208. The network 202, is, for example, the Internet, a local area network or a wide area network. The PDAs 204 and 206 represent wireless handheld devices that are able to communicate with the network 202 to interact with remote servers 210, also coupled to the network 202. The language processing server 208 is one particular remote server that the PDAs 204 and 206 are able to communicate with. The language processing server 208 includes a natural language processor for processing a voice or query input received over the network 208 from one of the PDAs 204 or 206. In addition, the language processing server 208 can also include a voice recognizer.

[0039] An example of the operation of the language processing system 200 is as follows, a user of the PDA 204 can input a voice query to the PDA 204. The voice input can then be digitized and transmitted by the PDA 204 to the language processing server 208 via the network 202. Once the voice

input is received at the language processing server 208, the language processing server 208 can perform processing on the voice query (e.g., voice recognition and/or natural language processing). In one embodiment, the language processing at the language processing server 208 can operate to interact with a knowledge base to understand the voice query in a natural language manner. The knowledge base can reside on the language processing server 208 or elsewhere on the network 202. After understanding the voice query, the language processing server 208 can return an indication of the meaning of the voice query to the PDA 204, such that the PDA 204 can operate in accordance with the voice query. The indication of the meaning of the voice input can cause the PDA 204 to perform various predetermined actions (e.g., perform an action, retrieve content, launch an application, etc.). For example the PDA 204 can retrieve information (stored within or remotely) that is being requested by the voice query.

[0040] In one embodiment, the voice query is asking for a particular content in a natural language manner. Hence, the language processing server 208 parses the voice query to understand the natural language nature of the query and interacts with the knowledge base, thereby understanding the question. Then, a database or other content resource can be accessed to retrieve responses to the understood query. The resulting responses, possibly with other appropriate resources or content, can then be delivered through the network 202 to the PDA 204.

[0041] Additional details on natural language processing can be found in U.S. Patent Nos.: 5,934,910; 5,884,302; and 5,836,771; all of which are hereby incorporated herein by reference. Still further, additional details on natural language processing can be found in U.S. Application No. 09/387,932, filed September 1, 1999, and U.S. Application No. 09/496,863, filed February 2, 2000, both of which are hereby incorporated herein by reference.

[0042] The examples given so far have used handheld personal digital assistant (PDA) devices as the example platform. This invention applies to PDAs, "smart" cellular phones, internet appliances, and other devices with limited input capabilities. The natural language processing can also be extended beyond voice input to other forms of input, such as Optical

Character Recognition (OCR) for scanned or faxed input. An another embodiment, the present invention is also applicable to voice inputs or OCR inputs for desktop computers.

[0043] The invention can be implemented in software and/or hardware. The invention can also be embodied as computer readable code on a computer readable medium. The computer readable medium is any data storage device that can store data which can thereafter be read by a computer system. Examples of the computer readable medium include read-only memory, random-access memory, CD-ROMs, DVDs, magnetic tape, optical data storage devices, and carrier waves. The computer readable medium can also be distributed over network-coupled computer systems so that the computer readable code is stored and executed in a distributed fashion.

[0044] Other embodiments of the invention will be apparent to those skilled in the art from a consideration of this specification or practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

What is claimed is: